

# The State-Dependent Semideterministic Broadcast Channel

Amos Lapidoth      Ligong Wang

## Abstract

We derive the capacity region of the state-dependent semideterministic broadcast channel with non-causal state-information at the transmitter. In this broadcast channel one of the outputs is a deterministic function of the channel input and the channel state, and the state is assumed to be known noncausally to the transmitter but not to the receivers.

## I. INTRODUCTION

We study the capacity region of the discrete, memoryless, state-dependent, semideterministic broadcast channel. Such a channel has a single transmitting node, two receiving nodes, and an internal state, all of which are assumed to take value in finite sets. One of the receiving nodes observes a symbol  $Y$  that is a deterministic function of the transmitted symbol  $x$  and the state  $S$

$$Y = f(x, S) \quad \text{with probability one,} \quad (1a)$$

and the other receiving node observes a symbol  $Z$  which is random: conditional on the input being  $x$  and the state being  $s$ , the probability that it be  $z$  is  $W(z|x, s)$

$$\Pr(Z = z|X = x, S = s) = W(z|x, s). \quad (1b)$$

The state sequence  $\mathbf{S}$  is assumed to be independent and identically distributed (IID) according to some law  $P_S(\cdot)$

$$\Pr(S = s) = P_S(s) \quad (1c)$$

A. Lapidoth is with the Signal and Information Processing Laboratory, ETH Zurich, Switzerland. E-mail: amos.lapidoth@ethz.ch.

L. Wang is with the Research Laboratory of Electronics, MIT, Cambridge, MA, USA. E-mail: wlg@mit.edu.

and to be revealed to the encoder in a noncausal way: all future values of the state are revealed to the transmitter before transmission begins.

The definition of the *capacity region* of this channel is analogous to that of a broadcast channel without a state [1], [2]. The main result of this paper is a single-letter characterization of the capacity region:

*Theorem 1:* The capacity region of the channel (1) when the states are known noncausally to the transmitter is the convex closure of the union of rate-pairs  $(R_1, R_2)$  satisfying

$$R_1 < H(Y|S) \quad (2a)$$

$$R_2 < I(U; Z) - I(U; S) \quad (2b)$$

$$R_1 + R_2 < H(Y|S) + I(U; Z) - I(U; S, Y) \quad (2c)$$

over all joint distribution on  $(X, Y, Z, S, U)$  whose marginal  $P_S$  is the given state distribution and under which, conditional on  $X$  and  $S$ , the channel outputs  $Y$  and  $Z$  are drawn according to the channel law (1) independently of  $U$ :

$$P_{XYZSU}(x, y, z, s, u) = P_S(s)P_{XU|S}(x, u|s)\mathbf{1}\{y = f(x, s)\}W(z|x, s). \quad (3)$$

Here  $\mathbf{1}\{\cdot\}$  denotes the indicator function.<sup>1</sup> Moreover, the capacity region remains the same even if the state sequence is revealed to the deterministic receiver, i.e., if we replace  $f(\cdot, \cdot)$  with the mapping  $(x, s) \mapsto (f(x, s), s)$ .

State-dependent broadcast channels were considered before [3], [4]. Steinberg [3] studied the *degraded* state-dependent broadcast channel with causal and with noncausal side-information at the transmitter. He derived the capacity region for the causal case, but for the noncausal case his outer and inner bounds do not coincide. Steinberg and Shamai [4] then derived an inner bound for general (not necessarily degraded) state-dependent broadcast channels with noncausal side-information. This inner bound is based on Marton's inner bound for broadcast channels without states [5] and on Gel'fand-Pinsker coding [6]. In fact, the direct part of our Theorem 1 can be deduced from [4] with a proper choice of the auxiliary random variables (see Section III-A). However, capacity regions of most state-dependent broadcast channels are still unknown.

<sup>1</sup>The value of  $\mathbf{1}\{\text{statement}\}$  is 1 if the statement is true and is 0 otherwise.

Much work has been done on broadcast channels *without states* [7]. Our work can be considered as an extension of previous works on deterministic broadcast channels (solved by Gel'fand, Marton and Pinsker [8]–[10]) and on semideterministic broadcast channels (solved by Gel'fand and Pinsker [11]). These results can be found in [2].

In the rest of this paper we discuss some special cases of Theorem 1 (Section II) and then prove the direct and converse parts of Theorem 1 in Sections III and IV.

## II. SIMPLE SPECIAL CASES

In this section we discuss two special cases of Theorem 1 with self-contained proofs which are simpler than the proof of Theorem 1. The first special case is a single-user deterministic channel described by (1a), where the states are known noncausally to the transmitter. In this case, Theorem 1 reduces to the following:

*Theorem 2:* The capacity of the single-user deterministic channel (1a) when the states are known noncausally to the transmitter is

$$C = \max_{P_{X|S}} H(Y|S) \quad (4)$$

$$= \sum_s P_S(s) \log |\{y: \exists x, y = f(x, s)\}|, \quad (5)$$

where the conditional entropy on the right-hand side (RHS) of (4) is computed for the joint distribution  $P_{XYS}(x, y, s) = P_S(s)P_{X|S}(x|s)\mathbf{1}\{y = f(x, s)\}$ . Furthermore, the capacity is unchanged when the state sequence is also revealed to the receiver.

*Proof:* To prove the direct part, we use the formula of Gel'fand and Pinsker [6]:

$$C = \max_{U \circ - (X, S) \circ - Y} I(U; Y) - I(U; S), \quad (6)$$

where we choose the auxiliary random variable  $U$  to be  $Y$ . Note that because  $Y$  is a deterministic function of  $X$  and  $S$ , this choice is valid in the sense that  $U \circ - (X, S) \circ - Y$  forms a Markov chain. Consequently,

$$C \geq I(Y; Y) - I(Y; S) \quad (7)$$

$$= H(Y) - (H(Y) - H(Y|S)) \quad (8)$$

$$= H(Y|S), \quad (9)$$

which, when optimized over  $P_{X|S}$ , proves the direct part.

To prove the converse, we first note that the capacity  $C$  is upper-bounded by the capacity  $C_{\text{both}}$  corresponding to the case where the state sequence is known to both the transmitter and the receiver, a capacity which is given by

$$C_{\text{both}} = \max_{P_{X|S}} I(X; Y|S) \quad (10)$$

$$= \max_{P_{X|S}} H(Y|S) - H(Y|X, S) \quad (11)$$

$$= \max_{P_{X|S}} H(Y|S). \quad (12)$$

This establishes (4). And from (4) follows (5) by noting that entropy is maximized by the uniform distribution.  $\blacksquare$

The second special case we consider is the *fully* deterministic broadcast channel, where the symbols received at *both* receiving nodes are deterministic functions of the channel input and the channel state:

$$Y = f_1(x, s) \quad (13a)$$

$$Z = f_2(x, s). \quad (13b)$$

For this channel we have the following:

*Theorem 3:* The capacity region of the channel (13) when the state sequence is drawn IID  $P_S$  and is known noncausally to the transmitter is the convex closure of the union of rate-pairs  $(R_1, R_2)$  satisfying

$$R_1 \leq H(Y|S) \quad (14a)$$

$$R_2 \leq H(Z|S) \quad (14b)$$

$$R_1 + R_2 \leq H(Y, Z|S) \quad (14c)$$

where the union is over all choices of  $P_{X|S}$ , and where the conditional entropies are computed with respect to the joint distribution  $P_{XYZS}(x, y, z, s) = P_S(s)P_{X|S}(x|s)\mathbf{1}\{y = f_1(x, s)\}\mathbf{1}\{z = f_2(x, s)\}$ . Moreover, this is also the capacity region when the state sequence is also revealed to both receivers.

*Proof sketch.* The direct part is proved like the direct part of Theorem 1 and is omitted. But the converse is greatly simplified: Let the state sequence be known to the transmitter and to both

receivers. For the sum rate, we have

$$n(R_1 + R_2) = H(M_1, M_2) \quad (15)$$

$$\leq I(M_1, M_2; Y^n, Z^n, S^n) + n\epsilon_n \quad (16)$$

$$= I(M_1, M_2; Y^n, Z^n | S^n) + n\epsilon_n \quad (17)$$

$$= \sum_{i=1}^n I(M_1, M_2; Y_i, Z_i | S^n, Y^{i-1}, Z^{i-1}) + n\epsilon_n \quad (18)$$

$$= \sum_{i=1}^n H(Y_i, Z_i | S^n, Y^{i-1}, Z^{i-1}) + n\epsilon_n \quad (19)$$

$$\leq \sum_{i=1}^n H(Y_i, Z_i | S_i) + n\epsilon_n, \quad (20)$$

where  $\epsilon_n$  is a function of  $n$  which decays to zero as  $n$  goes to infinity. Here, (16) follows from Fano's Inequality; (17) because  $S^n$  is independent of  $(M_1, M_2)$ ; (18) from the chain rule; (19) by expanding mutual information and recalling that the outputs are deterministic functions of the state and input; and (20) because conditioning cannot increase entropy. By similarly bounding  $R_1$  and  $R_2$  and letting  $n$  tend to infinity, we obtain the converse part of Theorem 3.  $\square$

### III. DIRECT PART

In this section we prove the direct part of Theorem 1. One way to do this is to use [4, Theorem 1] with a judicious choice of the auxiliary random variables, as we propose in Section III-A. For completeness and simplicity, we also provide a self-contained proof in Section III-B.

#### A. Proof based on [4]

It was shown in [4, Theorem 1] that the capacity region of any (not necessarily semideterministic) state-dependent broadcast channel with noncausal side-information at the transmitter contains the convex closure of the union of rate-pairs  $(R_1, R_2)$  satisfying

$$R_1 \leq I(U_0, U_1; Y) - I(U_0, U_1, S) \quad (21a)$$

$$R_2 \leq I(U_0, U_2; Z) - I(U_0, U_2, S) \quad (21b)$$

$$\begin{aligned} R_1 + R_2 \leq & -[\max\{I(U_0; Y), I(U_0; Z)\} - I(U_0; S)]^+ \\ & + I(U_0, U_1; Y) - I(U_0, U_1, S) \\ & + I(U_0, U_2; Z) - I(U_0, U_2, S) - I(U_1, U_2 | U_0, S), \end{aligned} \quad (21c)$$

where the union is over all joint distributions of  $(X, Y, Z, S, U_0, U_1, U_2)$  whose marginal is  $P_S$ ; that satisfy the Markov condition

$$(U_0, U_1, U_2) \text{---} (X, S) \text{---} (Y, Z); \quad (22)$$

and under which the conditional law of  $(Y, Z)$  given  $(X, S)$  is that of the given channel.

For the semideterministic channel, we choose the auxiliary random variables in (21) as follows:

$$U_0 = 0 \quad (\text{deterministic}) \quad (23a)$$

$$U_1 = Y \quad (23b)$$

$$U_2 = U. \quad (23c)$$

Note that the Markov condition (22) is satisfied because  $Y$  is a deterministic function of  $(X, S)$  and because in Theorem 1 we restrict  $U$  to be such that  $U \text{---} (X, S) \text{---} (Y, Z)$ . With this choice of  $U_0, U_1$ , and  $U_2$ , (21) reduces to (2).

### B. Self-Contained Proof

We next provide a self-contained proof of the direct part of Theorem 1. Like [4, Theorem 1], our proof is based on Marton's inner bound for general broadcast channels [5], [12] and on Gel'fand-Pinsker coding [6].

First note that the joint distribution (3) can also be written as

$$P_{XYZSU}(x, y, z, s, u) = P_S(s)P_{YU|S}(y, u|s)P_{X|YSU}(x|y, s, u)W(z|x, s) \quad (24)$$

with the additional requirement that

$$y = f(x, s) \quad \text{with probability one.} \quad (25)$$

Further note that, when  $P_{YSU}$  is fixed, all the terms on the RHS of (2) are fixed, except for  $I(U; Z)$ , which is convex in  $P_{X|YS}$  when  $P_{YSU}$  is held fixed. Since  $I(U; Z)$  only appears with a positive sign on the RHS of (2), it follows that the union over all joint distributions of the form (2) can be replaced by a union only over those where  $x$  is a deterministic function of  $(y, u, s)$ , i.e., of the form

$$P_{XYZSU}(x, y, z, s, u) = P_S(s)P_{YU|S}(y, u|s)\mathbf{1}\{x = g(y, u, s)\}W(z|x, s) \quad (26)$$

for some  $g: (y, u, s) \mapsto x$  (and subject to (25)). We shall thus only establish the achievability of rate pairs that satisfy (2) for some distribution of the form (26).

Choose a stochastic kernel  $P_{YU|S}$  and a mapping  $g: (y, u, s) \mapsto x$  which, combined with  $P_S$  and the channel law, determines the joint distribution (26) for which (25) is satisfied. For a given block-length  $n$ , we generate a random code as follows:

**Codebook:** Generate  $2^{nR_1}$   $y$ -bins, each containing  $2^{n\tilde{R}_1}$   $y$ -tuples where the  $l_1$ -th  $y$ -tuple in the  $m_1$ -th bin

$$\mathbf{y}(m_1, l_1), \quad m_1 \in \{1, \dots, 2^{nR_1}\}, \quad l_1 \in \{1, \dots, 2^{n\tilde{R}_1}\}$$

is generated IID according to  $P_Y$  (the  $Y$ -marginal of (26)) independently of the other  $y$ -tuples. Additionally, generate  $2^{nR_2}$   $u$ -bins, each containing  $2^{n\tilde{R}_2}$   $u$ -tuples, where the  $l_2$ -th  $u$ -tuple in the  $m_2$ -th  $u$ -bin

$$\mathbf{u}(m_2, l_2), \quad m_2 \in \{1, \dots, 2^{nR_2}\}, \quad l_2 \in \{1, \dots, 2^{n\tilde{R}_2}\}$$

is drawn IID according to  $P_U$  (the  $U$ -marginal of (26)) independently of the other  $u$ -tuples and of the  $y$ -tuples.

**Encoder:** To send message  $m_1 \in \{1, \dots, 2^{R_1}\}$  to Receiver 1 and message  $m_2 \in \{1, \dots, 2^{R_2}\}$  to Receiver 2, look for a  $y$ -tuple  $\mathbf{y}(m_1, l_1)$  in  $y$ -bin  $m_1$  and a  $u$ -tuple  $\mathbf{u}(m_2, l_2)$  in  $u$ -bin  $m_2$  such that  $(\mathbf{y}(m_1, l_1), \mathbf{u}(m_2, l_2))$  is jointly typical with the state sequence  $\mathbf{s}$ . If such a pair can be found, send

$$\mathbf{x} = g(\mathbf{y}(m_1, l_1), \mathbf{u}(m_2, l_2), \mathbf{s}), \quad (27)$$

where in the above  $g(\mathbf{y}, \mathbf{u}, \mathbf{s})$  denotes the application of the function  $g(y, u, s)$  componentwise. In this case the sequence received by the deterministic receiver will be  $\mathbf{y}(m_1, l_1)$ . Otherwise send an arbitrary codeword.

**Decoder 1:** Try to find the *unique*  $y$ -bin, say  $m'_1$ , that contains the received sequence  $\mathbf{y}$  and output its number  $m'_1$ . If there is more than one such bin, declare an error.

**Decoder 2:** Try to find the *unique*  $u$ -tuple  $\mathbf{u}(m'_2, l'_2)$  that is jointly typical with the received sequence  $\mathbf{z}$  and output its bin number  $m'_2$ . If more than one or no such  $\mathbf{u}$  can be found, declare an error.

We next analyze the error probability of the above coding scheme. There are three types of errors:

**Decoder 1 errs.** This happens only if there is more than one bin that contains the received  $\mathbf{y}$ . This probability tends to zero as  $n$  tends to infinity provided

$$R_1 + \tilde{R}_1 < H(Y). \quad (28)$$

**Decoder 2 errs.** This happens if either the  $u$ -tuple  $\mathbf{u}(m_2, l_2)$  is not jointly typical with the received  $z$ -tuple, or if a different  $u$ -tuple happens to be jointly typical with the received  $z$ -tuple. The probability of the former case tends to zero as  $n$  tends to infinity by the Markov Lemma [2]. The probability of the latter case tends to zero as  $n$  tends to infinity provided

$$R_2 + \tilde{R}_2 < I(U; Z). \quad (29)$$

**Encoder errs.** This happens only if there is no  $(l_1, l_2) \in \{1, \dots, 2^{n\tilde{R}_1}\} \times \{1, \dots, 2^{n\tilde{R}_2}\}$  such that  $(\mathbf{y}(m_1, l_1), \mathbf{u}(m_2, l_2))$  is jointly typical with the state sequence  $\mathbf{s}$ . To bound this error probability, we generalize the Mutual Covering Lemma (see [12]). Without loss of generality, assume that the messages to be transmitted are both 1. Define

$$A_{l_1, l_2} \triangleq \mathbf{1} \{(\mathbf{y}(1, l_1), \mathbf{u}(1, l_2)) \text{ is jointly typical with } \mathbf{s}\} \quad (30)$$

and

$$A \triangleq \sum_{l_1=1}^{2^{n\tilde{R}_1}} \sum_{l_2=1}^{2^{n\tilde{R}_2}} A_{l_1, l_2}. \quad (31)$$

The Encoder errs if, and only if,

$$A = 0. \quad (32)$$

By Chebyshev's Inequality

$$\Pr[A = 0] \leq \Pr[|A - \mathbf{E}[A]| \geq \mathbf{E}[A]] \leq \frac{\text{Var}(A)}{\mathbf{E}[A]^2}. \quad (33)$$

To bound  $\mathbf{E}[A]$ , note that for any  $l_1$  and  $l_2$  the vectors  $\mathbf{y}(1, l_1)$ ,  $\mathbf{u}(1, l_2)$  and  $\mathbf{s}$  are generated independently according to their marginal distributions. Hence, for large enough  $n$ , the probability that they are jointly typical is lower-bounded by

$$\Pr[A_{l_1, l_2} = 1] \geq 2^{-n(H(Y) + H(U) + H(S) - H(Y, U, S) + 4\epsilon)} \quad (34)$$



so

$$\mathbf{E}[A] = \sum_{l_1, l_2} \mathbf{E}[A_{l_1, l_2}] \quad (35)$$

$$= \sum_{l_1, l_2} \Pr[A_{l_1, l_2} = 1] \quad (36)$$

$$\geq 2^n \left( \tilde{R}_1 + \tilde{R}_2 - H(Y) - H(U) - H(S) + H(Y, U, S) - 4\epsilon \right). \quad (37)$$

To bound  $\text{Var}(A)$  we first write it as

$$\begin{aligned} \text{Var}(A) &= \sum_{l_1, l_2} \text{Var}(A_{l_1, l_2}) + \sum_{\substack{l_1, l_2 \\ l'_1 \neq l_1}} \text{Cov}[A_{l_1, l_2}, A_{l'_1, l_2}] \\ &\quad + \sum_{\substack{l_1, l_2 \\ l'_2 \neq l_2}} \text{Cov}[A_{l_1, l_2}, A_{l_1, l'_2}] + \sum_{\substack{l_1, l_2 \\ l'_1 \neq l_1 \\ l'_2 \neq l_2}} \text{Cov}[A_{l_1, l_2}, A_{l'_1, l'_2}], \end{aligned} \quad (38)$$

and then bound the four terms on the RHS of (38) separately. For the first term, note that

$$\text{Var}(A_{l_1, l_2}) = \mathbf{E}[A_{l_1, l_2}^2] - \mathbf{E}[A_{l_1, l_2}]^2 \quad (39)$$

$$\leq \mathbf{E}[A_{l_1, l_2}^2] \quad (40)$$

$$= \Pr[A_{l_1, l_2} = 1], \quad (41)$$

so

$$\sum_{l_1, l_2} \text{Var}(A_{l_1, l_2}) \leq \sum_{l_1, l_2} \Pr[A_{l_1, l_2} = 1] \quad (42)$$

$$= \mathbf{E}[A]. \quad (43)$$

For the second term on the RHS of (38) we have

$$\text{Cov}[A_{l_1, l_2}, A_{l'_1, l_2}] \leq \mathbf{E}[A_{l_1, l_2} A_{l'_1, l_2}] \quad (44)$$

$$= \Pr[A_{l_1, l_2} = A_{l'_1, l_2} = 1]. \quad (45)$$

The probability that  $\mathbf{u}(1, l_2)$  and  $\mathbf{s}$  are jointly typical is upper-bounded by  $2^{-n(I(U; S) - 3\epsilon)}$ . Given that  $(\mathbf{u}(1, l_2), \mathbf{s})$  are jointly typical, the sequences  $\mathbf{y}(1, l_1)$  and  $\mathbf{y}(1, l'_1)$  are independent, and each of them is jointly typical with  $(\mathbf{u}(1, l_2), \mathbf{s})$  with probability at most  $2^{-n(I(Y; U, S) - 3\epsilon)}$ . Hence

$$\text{Cov}[A_{l_1, l_2}, A_{l'_1, l_2}] \leq 2^{-n(I(U; S) - 3\epsilon)} \cdot \left( 2^{-n(I(Y; U, S) - 3\epsilon)} \right)^2 \quad (46)$$

$$= 2^{-n(I(U; S) + 2I(Y; U, S) - 9\epsilon)}, \quad (47)$$

so

$$\sum_{\substack{l_1, l_2 \\ l'_1 \neq l_1}} \text{Cov}[A_{l_1, l_2}, A_{l'_1, l_2}] \leq 2^{n(2\tilde{R}_1 + \tilde{R}_2 - I(U; S) - 2I(Y; U, S) + 9\epsilon)}. \quad (48)$$

Similarly for the third term on the RHS of (38) we have

$$\sum_{\substack{l_1, l_2 \\ l'_2 \neq l_2}} \text{Cov}[A_{l_1, l_2}, A_{l_1, l'_2}] \leq 2^{n(\tilde{R}_1 + 2\tilde{R}_2 - I(Y; S) - 2I(U; Y, S) + 9\epsilon)}. \quad (49)$$

We next analyze the fourth term on the RHS of (38). When  $l'_1 \neq l_1$  and  $l'_2 \neq l_2$  the random variables  $A_{l_1, l_2}$  and  $A_{l'_1, l'_2}$  are independent and hence uncorrelated. Consequently

$$\sum_{\substack{l_1, l_2 \\ l'_1 \neq l_1 \\ l'_2 \neq l_2}} \text{Cov}[T_{l_1, l_2}, T_{l'_1, l'_2}] = 0. \quad (50)$$

Summarizing (33), (37), (38), (43), (48), (49) and (50) we conclude that the probability that  $A$  is zero, i.e., that the encoder errs, tends to zero as  $n$  tends to infinity provided that  $\epsilon$  tends to zero and

$$\tilde{R}_1 > I(Y; S) \quad (51a)$$

$$\tilde{R}_2 > I(U; S) \quad (51b)$$

$$\tilde{R}_1 + \tilde{R}_2 > H(Y) + H(U) + H(S) - H(Y, U, S). \quad (51c)$$

Summarizing (28), (29) and (51) we conclude that the above coding scheme has vanishing error probability as  $n$  tends to infinity for all  $(R_1, R_2)$  satisfying (2). By time-sharing we further achieve the convex hull of all rate-pairs satisfying (2) for joint distributions of the form (26). This concludes the proof of the direct part of Theorem 1.

#### IV. CONVERSE PART

Our proof of the converse part of Theorem 1 employs ideas from Nair and El Gamal's outer bound [13] and of Gel'fand and Pinsker's converse for the state-dependent single-user channel [6], but it also contains some new elements.

We shall show that, even if the state sequence  $\mathbf{S}$  is revealed to the deterministic receiver (which observes  $\mathbf{Y}$ ), any achievable rate-pair must be in the convex closure of the union of rate-pairs

satisfying (2). Given any code of block-length  $n$ , we first derive a bound on  $R_1$ :

$$nR_1 = H(M_1) \quad (52)$$

$$\leq I(M_1; Y^n, S^n) + n\epsilon_n \quad (53)$$

$$= I(M_1; Y^n | S^n) + n\epsilon_n \quad (54)$$

$$= \sum_{i=1}^n I(M_1; Y_i | Y^{i-1}, S^n) + n\epsilon_n \quad (55)$$

$$\leq \sum_{i=1}^n H(Y_i | Y^{i-1}, S^n) + n\epsilon_n \quad (56)$$

$$\leq \sum_{i=1}^n H(Y_i | S_i) + n\epsilon_n, \quad (57)$$

where  $\epsilon_n$  is a function of  $n$  which decays to zero as  $n$  goes to infinity. Here, (53) follows from Fano's Inequality; (54) because  $M_1$  and  $S^n$  are independent; (55) from the chain rule; (56) by dropping negative terms; and (57) because conditioning cannot increase entropy.

We next bound  $R_2$  as in [6]:

$$nR_2 = H(M_2) \quad (58)$$

$$\leq I(M_2; Z^n) + n\epsilon_n \quad (59)$$

$$= \sum_{i=1}^n I(M_2; Z_i | Z^{i-1}) + n\epsilon_n \quad (60)$$

$$= \sum_{i=1}^n I(M_2, S_{i+1}^n; Z_i | Z^{i-1}) - \sum_{i=1}^n I(S_{i+1}^n; Z_i | M_2, Z^{i-1}) + n\epsilon_n \quad (61)$$

$$= \sum_{i=1}^n I(M_2, S_{i+1}^n; Z_i | Z^{i-1}) - \sum_{i=1}^n I(Z^{i-1}; S_i | M_2, S_{i+1}^n) + n\epsilon_n \quad (62)$$

$$= \sum_{i=1}^n I(M_2, S_{i+1}^n; Z_i | Z^{i-1}) - \sum_{i=1}^n I(M_2, Z^{i-1}, S_{i+1}^n; S_i) + n\epsilon_n \quad (63)$$

$$\leq \sum_{i=1}^n I(M_2, Z^{i-1}, S_{i+1}^n; Z_i) - \sum_{i=1}^n I(M_2, Z^{i-1}, S_{i+1}^n; S_i) + n\epsilon_n \quad (64)$$

$$= \sum_{i=1}^n I(V_i; Z_i) - I(V_i; S_i) + n\epsilon_n. \quad (65)$$

Here, (59) follows from Fano's Inequality; (60) and (61) from the chain rule; (62) from Csiszár's

Identity [14]

$$\sum_{i=1}^n I(C_{i+1}^n; D_i | D^{i-1}) = \sum_{i=1}^n I(D^{i-1}; C_i | C_{i+1}^n); \quad (66)$$

(63) because  $S_i$  and  $(M_2, S^{i-1})$  are independent; (64) from the chain rule and by dropping negative terms; and (65) by defining the auxiliary random variables

$$V_i \triangleq (M_2, Z^{i-1}, S_{i+1}^n), \quad i \in \{1, \dots, n\}. \quad (67)$$

We next bound the sum rate  $R_1 + R_2$ :

$$n(R_1 + R_2) = H(M_1, M_2) \quad (68)$$

$$= H(M_2) + H(M_1 | M_2) \quad (69)$$

$$\leq I(M_2; Z^n) + I(M_1; Y^n, S^n | M_2) + n\epsilon_n, \quad (70)$$

where the last step follows from Fano's Inequality. Of the two mutual informations on the RHS of (70) we first bound  $I(M_2; Z^n)$ :

$$I(M_2; Z^n) = \sum_{i=1}^n I(M_2; Z_i | Z^{i-1}) \quad (71)$$

$$\leq \sum_{i=1}^n I(M_2, Z^{i-1}; Z_i) \quad (72)$$

$$\begin{aligned} &= \sum_{i=1}^n I(M_2, Z^{i-1}, S_{i+1}^n, Y_{i+1}^n; Z_i) \\ &\quad - \sum_{i=1}^n I(S_{i+1}^n, Y_{i+1}^n; Z_i | M_2, Z^{i-1}) \end{aligned} \quad (73)$$

$$\begin{aligned} &= \sum_{i=1}^n I(M_2, Z^{i-1}, S_{i+1}^n, Y_{i+1}^n; Z_i) \\ &\quad - \sum_{i=1}^n I(Z^{i-1}; S_i, Y_i | M_2, S_{i+1}^n, Y_{i+1}^n) \end{aligned} \quad (74)$$

$$\begin{aligned} &= \sum_{i=1}^n I(M_2, Z^{i-1}, S_{i+1}^n, Y_{i+1}^n; Z_i) \\ &\quad - \sum_{i=1}^n I(M_2, Z^{i-1}, S_{i+1}^n, Y_{i+1}^n; S_i, Y_i) \\ &\quad + \sum_{i=1}^n I(M_2, S_{i+1}^n, Y_{i+1}^n; S_i, Y_i). \end{aligned} \quad (75)$$

Here, (71), (72) and (73) follow from the chain rule; (74) by applying Csiszár's Identity (66) between  $(S^n, Y^n)$  and  $Z^n$ ; and (75) again from the chain rule.

We next study the sum of the last term on the RHS of (75) and the second mutual information on the RHS of (70):

$$\begin{aligned} & \sum_{i=1}^n I(M_2, S_{i+1}^n, Y_{i+1}^n; S_i, Y_i) + I(M_1; Y^n, S^n | M_2) \\ &= \sum_{i=1}^n I(M_2, S_{i+1}^n, Y_{i+1}^n; S_i, Y_i) \\ & \quad + \sum_{i=1}^n I(M_1; S_i, Y_i | M_2, S_{i+1}^n, Y_{i+1}^n) \end{aligned} \quad (76)$$

$$= \sum_{i=1}^n I(M_1, M_2, S_{i+1}^n, Y_{i+1}^n; S_i, Y_i) \quad (77)$$

$$\begin{aligned} &= \sum_{i=1}^n I(M_1, M_2, S_{i+1}^n, Y_{i+1}^n; S_i, Y_i) \\ & \quad + \sum_{i=1}^n I(S^{i-1}; S_i, Y_i | M_1, M_2, S_{i+1}^n, Y_{i+1}^n) \\ & \quad - \sum_{i=1}^n I(S_{i+1}^n, Y_{i+1}^n; S_i | M_1, M_2, S^{i-1}) \end{aligned} \quad (78)$$

$$\begin{aligned} &= \sum_{i=1}^n I(M_1, M_2, S^{i-1}, S_{i+1}^n, Y_{i+1}^n; S_i, Y_i) \\ & \quad - \sum_{i=1}^n I(S_{i+1}^n, Y_{i+1}^n; S_i | M_1, M_2, S^{i-1}) \end{aligned} \quad (79)$$

$$\begin{aligned} &= \sum_{i=1}^n I(M_1, M_2, S^{i-1}, S_{i+1}^n, Y_{i+1}^n; S_i, Y_i) \\ & \quad - \sum_{i=1}^n I(M_1, M_2, S^{i-1}, S_{i+1}^n, Y_{i+1}^n; S_i) \end{aligned} \quad (80)$$

$$= \sum_{i=1}^n I(M_1, M_2, S^{i-1}, S_{i+1}^n, Y_{i+1}^n; Y_i | S_i) \quad (81)$$

$$= \sum_{i=1}^n H(Y_i | S_i). \quad (82)$$

Here, (76) and (77) follow from the chain rule; (78) by applying Csiszár's Identity between

$(S^n, Y^n)$  and  $S^n$ , which yields

$$\sum_{i=1}^n I(S_{i+1}^n, Y_{i+1}^n; S_i | M_1, M_2, S^{i-1}) = \sum_{i=1}^n I(S^{i-1}; S_i, Y_i | M_1, M_2, S_{i+1}^n, Y_{i+1}^n); \quad (83)$$

(79) from the chain rule; (80) because  $S_i$  and  $(M_1, M_2, S^{i-1})$  are independent; (81) again from the chain rule; and (82) because, given  $(M_1, M_2, S^n)$ , the channel inputs  $X^n$  are determined by the encoder, and hence  $Y^n$  are also determined, so

$$H(Y_i | M_1, M_2, S^n, Y_{i+1}^n) = 0. \quad (84)$$

Combining (70), (75) and (82), using the definitions (67), and further defining

$$T_i \triangleq Y_{i+1}^n, \quad i \in \{1, \dots, n\}, \quad (85)$$

we obtain

$$\begin{aligned} n(R_1 + R_2) &\leq \sum_{i=1}^n I(V_i, T_i; Z_i) - \sum_{i=1}^n I(V_i, T_i; S_i, Y_i) \\ &\quad + \sum_{i=1}^n H(Y_i | S_i) + n\epsilon_n. \end{aligned} \quad (86)$$

Summarizing (57), (65) and (86) and letting  $n$  go to infinity we obtain that any achievable rate-pair  $(R_1, R_2)$  must be contained in the convex closure of the union of rate-pairs satisfying

$$R_1 < H(Y|S) \quad (87a)$$

$$R_2 < I(V; Z) - I(V; S) \quad (87b)$$

$$R_1 + R_2 < H(Y|S) + I(V, T; Z) - I(V, T; S, Y) \quad (87c)$$

where, given  $(X, S)$ , the outputs  $(Y, Z)$  are drawn according to the channel law (1) independently of the auxiliary random variables  $(V, T)$ . To prove the converse part of Theorem 1, it remains to replace  $V$  and  $T$  with a single auxiliary random variable. I.e., it remains to find an auxiliary random variable  $U$  such that

$$I(V; Z) - I(V; S) \leq I(U; Z) - I(U; S), \quad (88a)$$

$$\begin{aligned} H(Y|S) + I(V, T; Z) - I(V, T; S, Y) \\ \leq H(Y|S) + I(U; Z) - I(U; S, Y). \end{aligned} \quad (88b)$$

In fact, as we shall see, either choosing  $U$  to be  $V$  will satisfy (88) or else choosing it to be  $(V, T)$  will satisfy (88). If we choose  $U = V$ , then (88a) is satisfied with equality, and the requirement (88b) becomes

$$I(T; Z|V) - I(T; S, Y|V) \leq 0. \quad (89)$$

On the other hand, if we choose  $U = (V, T)$ , then (88b) is satisfied with equality, and the requirement (88a) becomes

$$I(T; Z|V) - I(T; S|V) \geq 0. \quad (90)$$

It remains to show that at least one of the requirements (89) and (90) must be satisfied: if it is (89), then we shall choose  $U$  as  $V$ , and if it is (90), then we shall choose  $U$  as  $(V, T)$ . To this end we note that for all random variables  $T, Z, V, S, Y$

$$I(T; Z|V) - I(T; S, Y|V) \leq I(T; Z|V) - I(T; S|V), \quad (91)$$

because the RHS minus the left-hand side equals  $I(T; Y|S, V)$  which is nonnegative. Therefore, *at least one of* (89) and (90) must be satisfied. We have thus shown that there must exist a  $U$  which satisfies both inequalities in (88), hence the bounds (87) can be relaxed to (2). This concludes the proof of the converse part of Theorem 1.

#### ACKNOWLEDGMENTS

L.W. acknowledges support from the US Air Force Office of Scientific Research under Grant No. FA9550-11-1-0183 and the National Science Foundation under Grant No. CCF-1017772.

#### REFERENCES

- [1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 1991.
- [2] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.
- [3] Y. Steinberg, "Coding for the degraded broadcast channel with random parameters, with causal and noncausal side information," *IEEE Trans. Inform. Theory*, vol. 51, no. 8, pp. 2867–2877, Aug. 2005.
- [4] Y. Steinberg and S. Shamai (Shitz), "Achievable rates for the broadcast channel with states known at the transmitter," in *Proc. IEEE Int. Symp. Inform. Theory*, Adelaide, Australia, Sept. 4–9, 2005, pp. 2184–2188.
- [5] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. 25, no. 3, pp. 306–311, May 1979.
- [6] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Prob. Contr. and Inform. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

- [7] T. M. Cover, “Comments on broadcast channels,” *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2524–2530, Oct. 1998.
- [8] S. I. Gel’fand, “Capacity of one broadcast channel,” *Problemy Peredachi Informatsii (Problems of Inform. Transm.)*, vol. 13, no. 3, pp. 106–108, July–Sept. 1977.
- [9] K. Marton, “The capacity region of deterministic broadcast channels,” *Trans. Int. Symp. Inform. Theory*, 1977.
- [10] M. S. Pinsker, “Capacity of noiseless broadcast channels,” *Problemy Peredachi Informatsii (Problems of Inform. Transm.)*, vol. 14, no. 2, pp. 28–34, Apr.–June 1978.
- [11] S. I. Gel’fand and M. S. Pinsker, “Capacity of a broadcast channel with one deterministic component,” *Problemy Peredachi Informatsii (Problems of Inform. Transm.)*, vol. 16, no. 1, pp. 17–25, Jan.–Mar. 1980.
- [12] A. El Gamal and E. C. van der Meulen, “A proof of Marton’s coding theorem for the discrete memoryless broadcast channel,” *IEEE Trans. Inform. Theory*, vol. 27, no. 1, pp. 120–122, Jan. 1981.
- [13] C. Nair and A. El Gamal, “An outer bound to the capacity region of the broadcast channel,” *IEEE Trans. Inform. Theory*, vol. 53, no. 1, pp. 50–55, Jan. 2007.
- [14] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1981.